www.bsc.es

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Reproducibility of EC-Earth

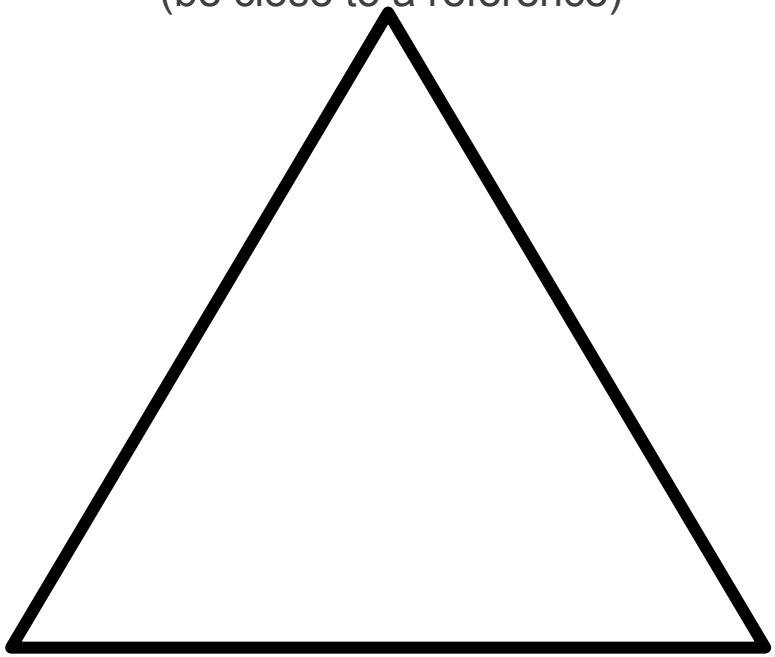
Towards a protocol for CMIP6

F. Massonnet, M. Ménégoz, M. Acosta, X. Yepes, O. Bellprat, M. Asif, E. Exarchou, K. Serraddell, O. Mula-Valls, V. Guemas, F. Doblas-Reyes

Reading, EC-Earth meeting, November 2016

Model development has the following objectives

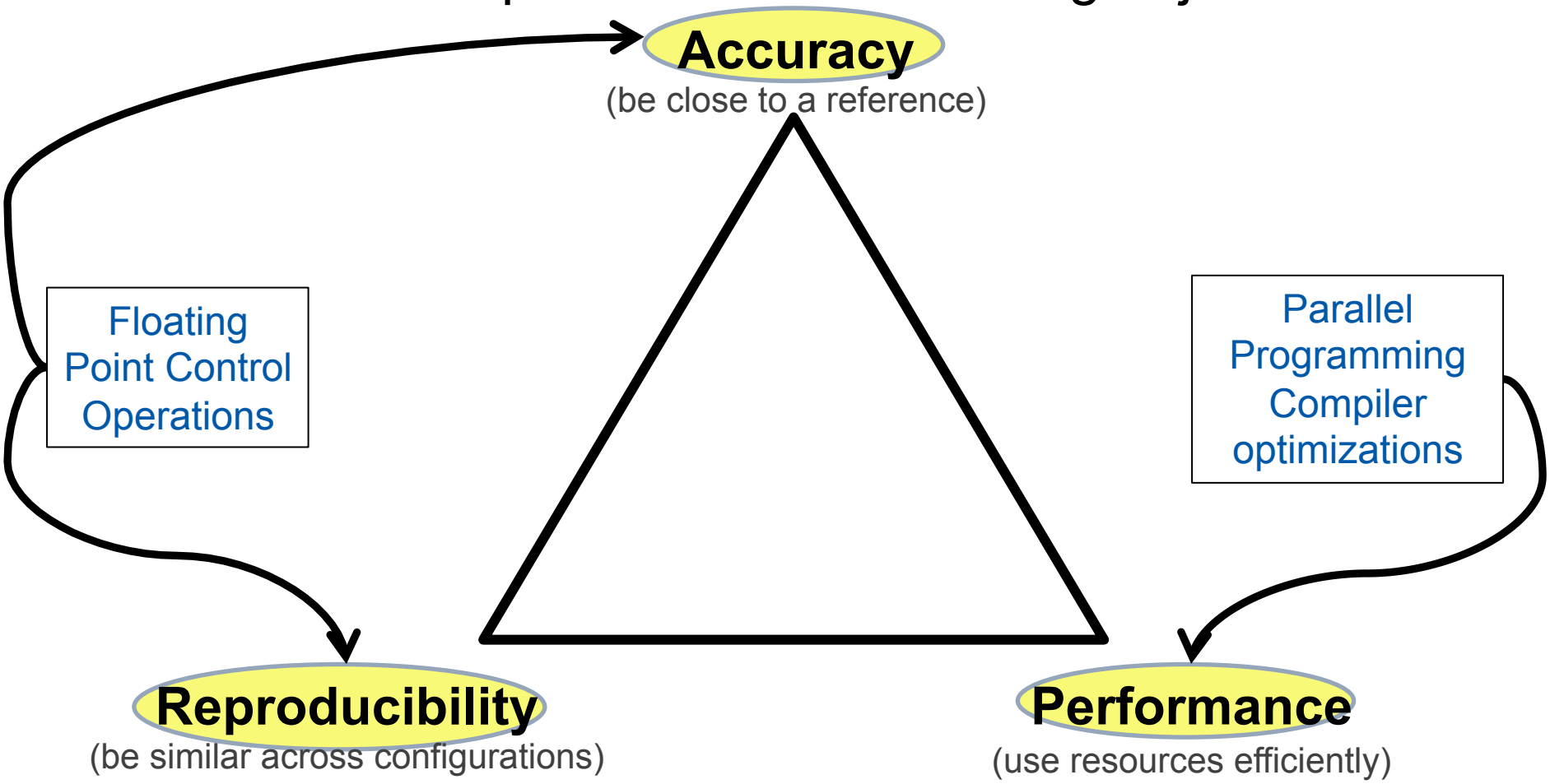
Accuracy
(be close to a reference)



Reproducibility
(be similar across configurations)

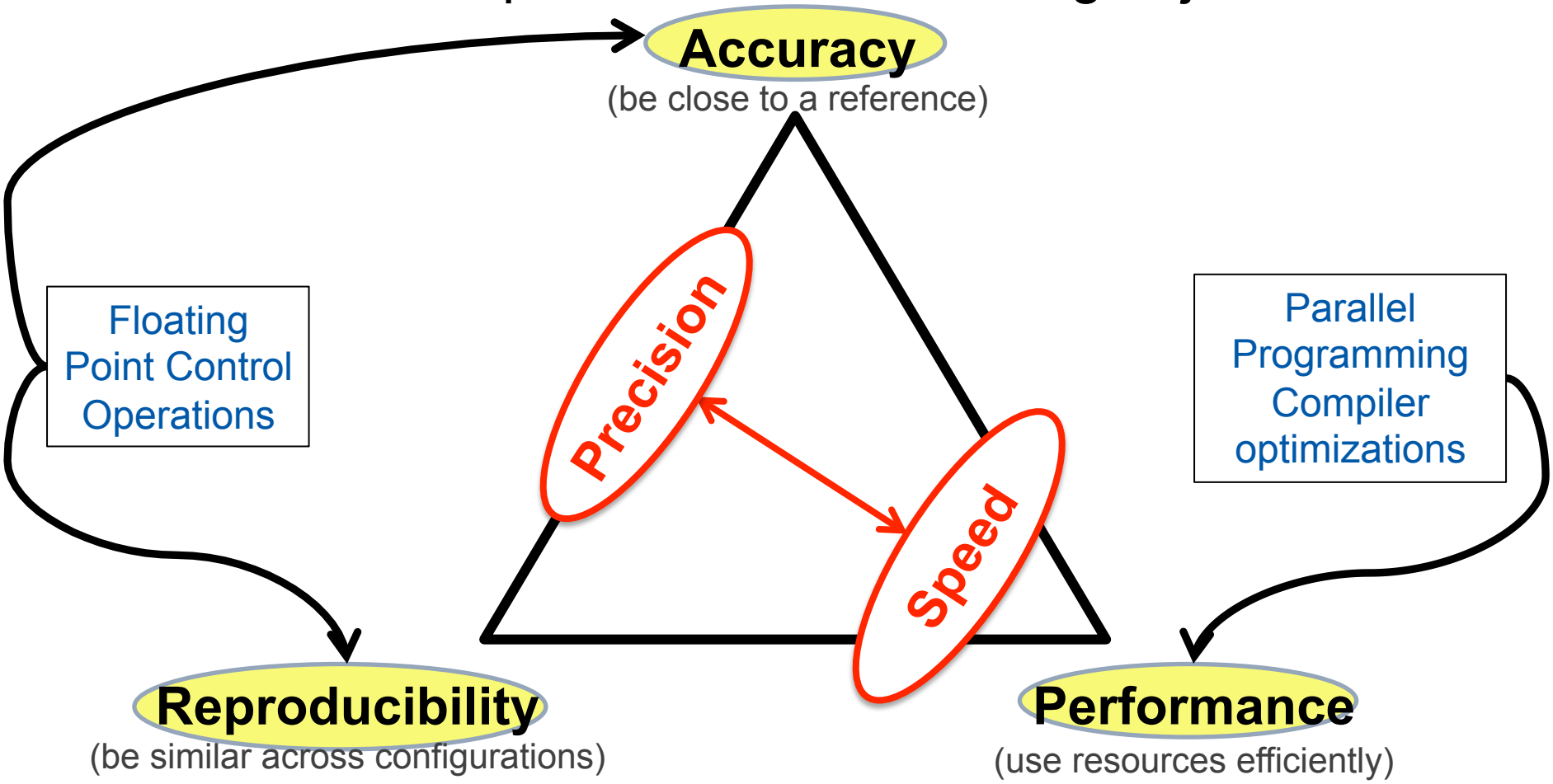
Performance
(use resources efficiently)

Model development has the following objectives



Compiler options let you control the tradeoffs among accuracy, reproducibility and performance.

Model development has the following objectives



Compiler options let you control the tradeoffs among accuracy, reproducibility and performance.

CMIP6 with EC-Earth

- CMIP6 start today, can we distribute the EC-Earth model on platforms that have very different characteristics?
- Are we underestimating the hardware uncertainty?
- Can we exchange restart files from center to center?

All these important (and equivalent) questions can only be answered if a strict protocol is developed, and applied every time a new model version is available. At BSC, we joined the IT team and the climate prediction group to address this question.

CMIP6 with EC-Earth

- CMIP6 start today, can we distribute the EC-Earth model on platforms that have very different characteristics?
- Are we underestimating the hardware uncertainty?
- Can we exchange restart files from center to center?

All these important (and equivalent) questions can only be answered if a strict protocol is developed, and applied every time a new model version is available. At BSC, we joined the IT team and the climate prediction group to address this question.

1. **The protocol**
2. **Application to EC-Earth 3.1 and 3.2**
3. **Recommendations**

1. The protocol

Defining a protocol for global climate simulations

20-yr long, 5-member, pre-industrial, coupled simulations

Defining a protocol for global climate simulations

20-yr long, 5-member, pre-industrial, coupled simulations

Allows to look at impact of machine on mean state/bias (not possible in the case of 1-yr simulations)

Defining a protocol for global climate simulations

20-yr long, 5-member, pre-industrial, coupled simulations

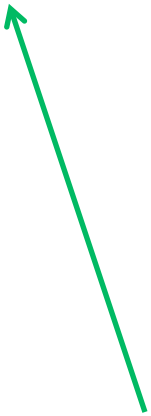
Allows to look at impact of machine on mean state/bias (not possible in the case of 1-yr simulations)

Allows to measure differences due to hardware as compared to internal variability

Defining a protocol for global climate simulations

20-yr long, 5-member, pre-industrial, coupled simulations

Allows to look at impact of machine on mean state/bias (not possible in the case of 1-yr simulations)

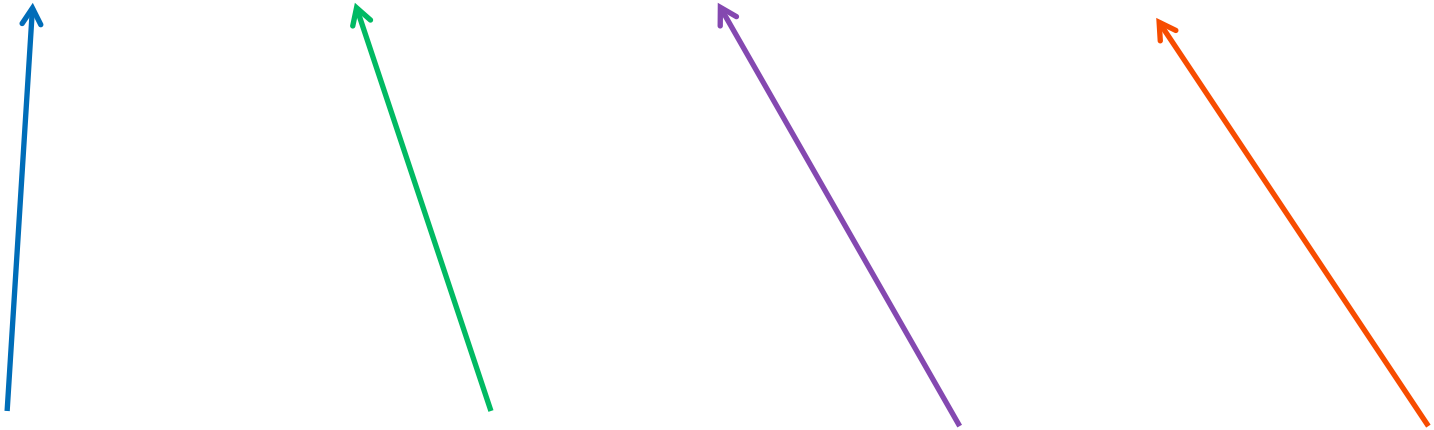


Allows to measure differences due to hardware as compared to internal variability

Working under stationary conditions removes possible dependence of hardware impact on the mean state

Defining a protocol for global climate simulations

20-yr long, 5-member, pre-industrial, coupled simulations



Allows to look at impact of machine on mean state/bias (not possible in the case of 1-yr simulations)

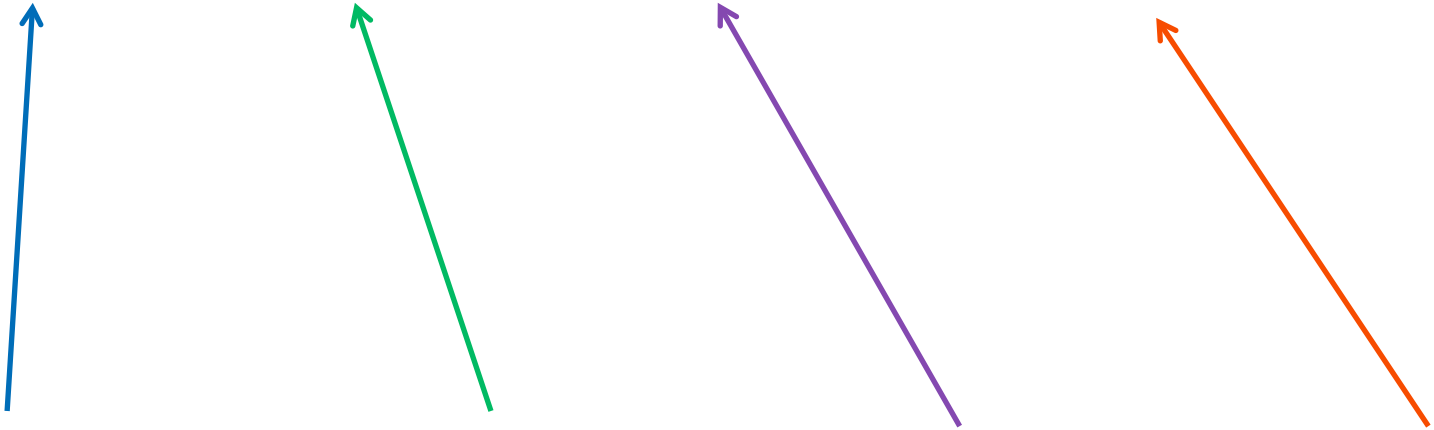
Allows to measure differences due to hardware as compared to internal variability

Working under stationary conditions removes possible dependence of hardware impact on the mean state

Addresses the problem from a global point-of-view; suitable to give recommendations for CMIP6

Defining a protocol for global climate simulations

20-yr long, 5-member, pre-industrial, coupled simulations



Allows to look at impact of machine on mean state/bias (not possible in the case of 1-yr simulations)

Allows to measure differences due to hardware as compared to internal variability

Working under stationary conditions removes possible dependence of hardware impact on the mean state

Addresses the problem from a global point-of-view; suitable to give recommendations for CMIP6

- Comparing experiences run on different platforms between themselves
- Computing their distance from a common reference (Reichler & Kim, 2008)

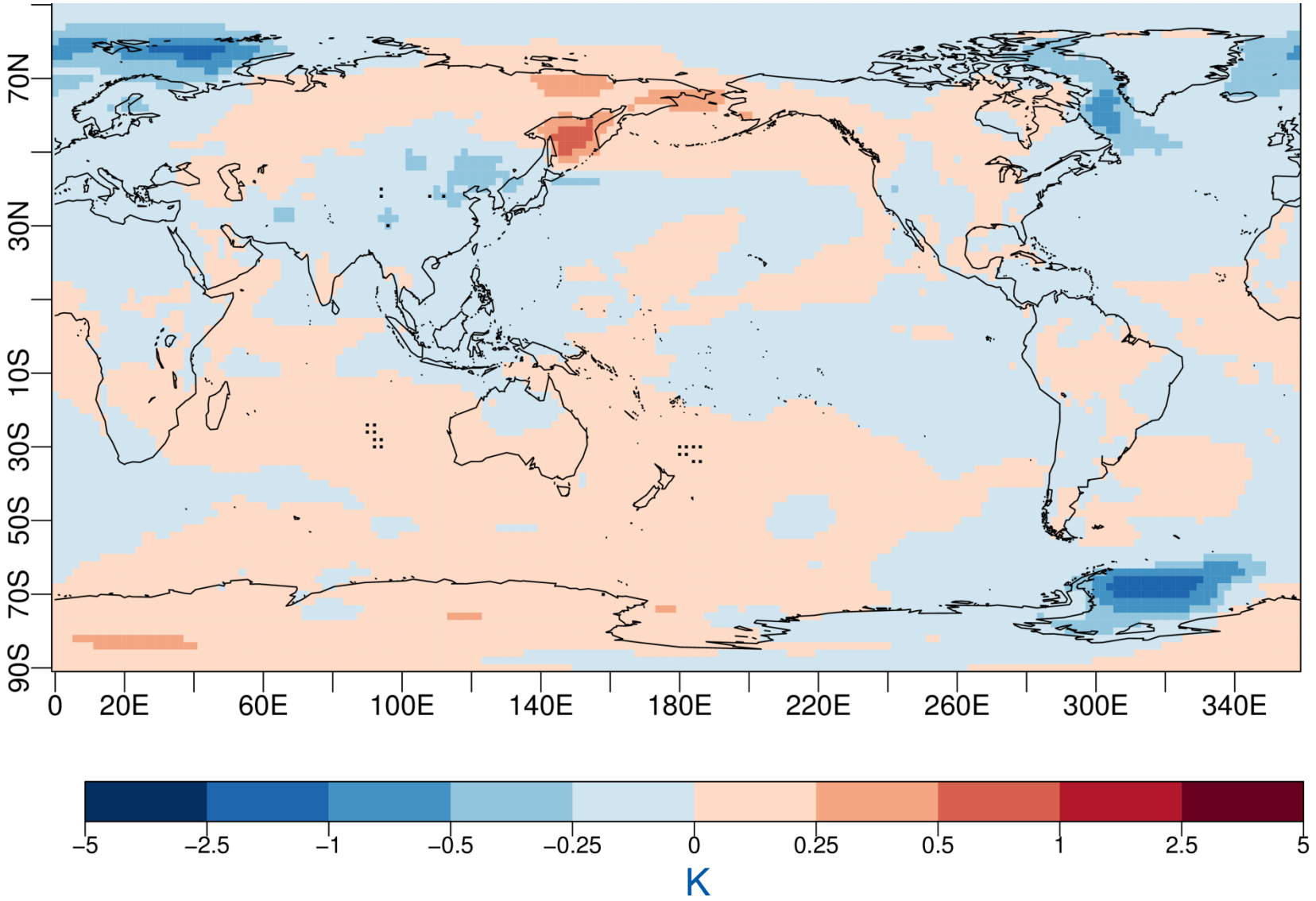
2. Application to EC-Earth

Application to EC-Earth 3.1

- Set of EC-Earth V3.1 experiments (IFS-NEMO-OASIS coupled).
- Same compilation options used on three different platforms (ECMWF, Marenostrum and Ithaca). Note:
  - Compiling with `-O2 -g -traceback -vec-report0:` model runs
  - Compiling with `-O2 -fp-model precise -fimf-arch-consistency=true -no-fma -g -traceback -vec-report0 -r8` model runs
  - Compiling with `-O2 -fp-model precise -fimf-arch-consistency=true -no-fma -g -traceback -vec-report0 -r8 -fpe0:` model cannot run (crashes after 3 time steps)
- The number of processors used is the same (72) except in sensitivity experiments labeled « HighProc » where 512 are used.
- The libraries (NetCDF, GRIBEX, GRIBAPI, etc.) are the default ones on each machine – they are from different versions and have not necessarily been compiled with the same options
- A Kolmogorov-Smirnov test is done to detect if differences between two hardware configurations are systematically greater than internal variability.

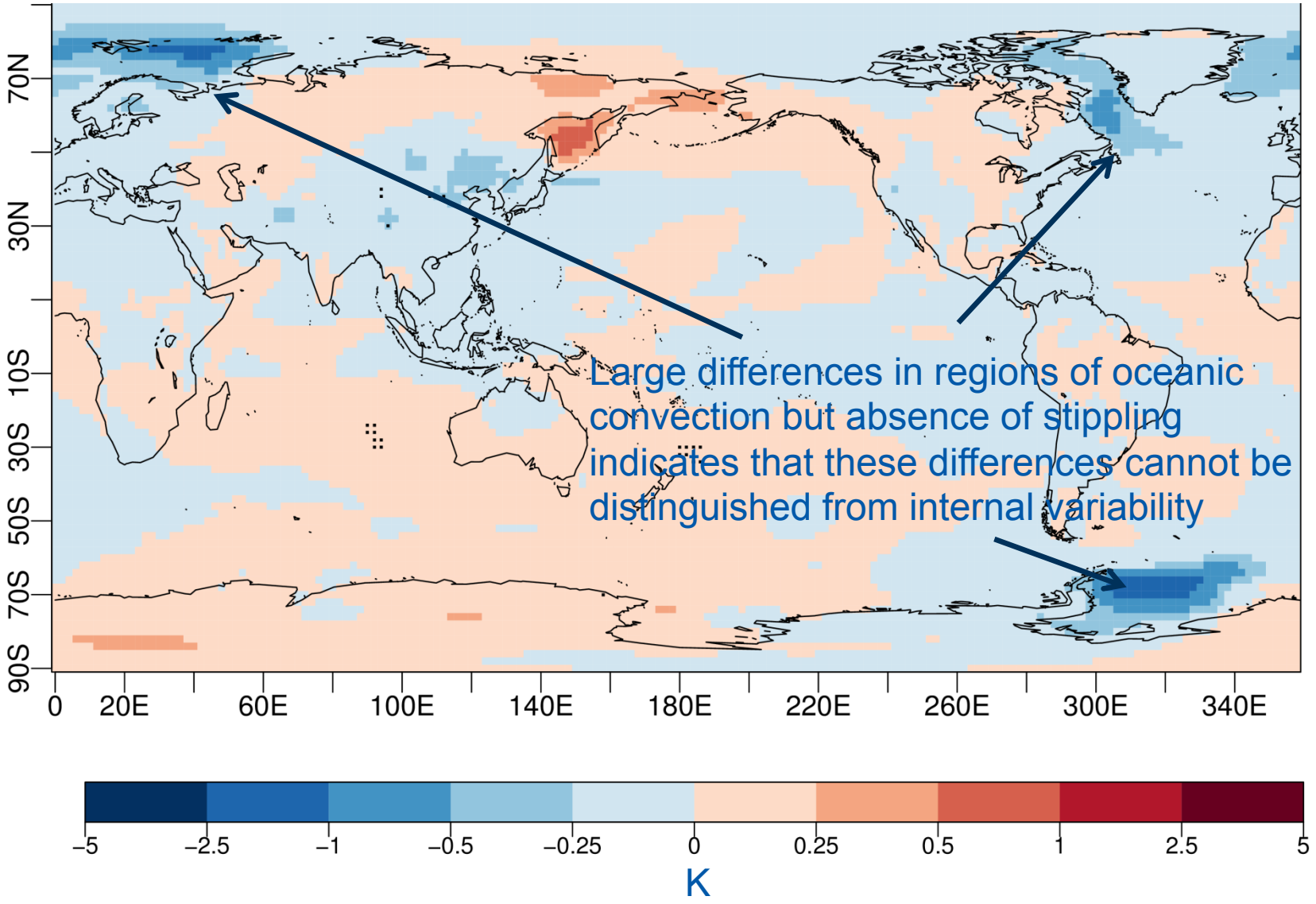
Changing the number of processors only does not affect the results

Difference of mean T2M for MN3-HighProc minus MN3. Stippling = significant at 5%



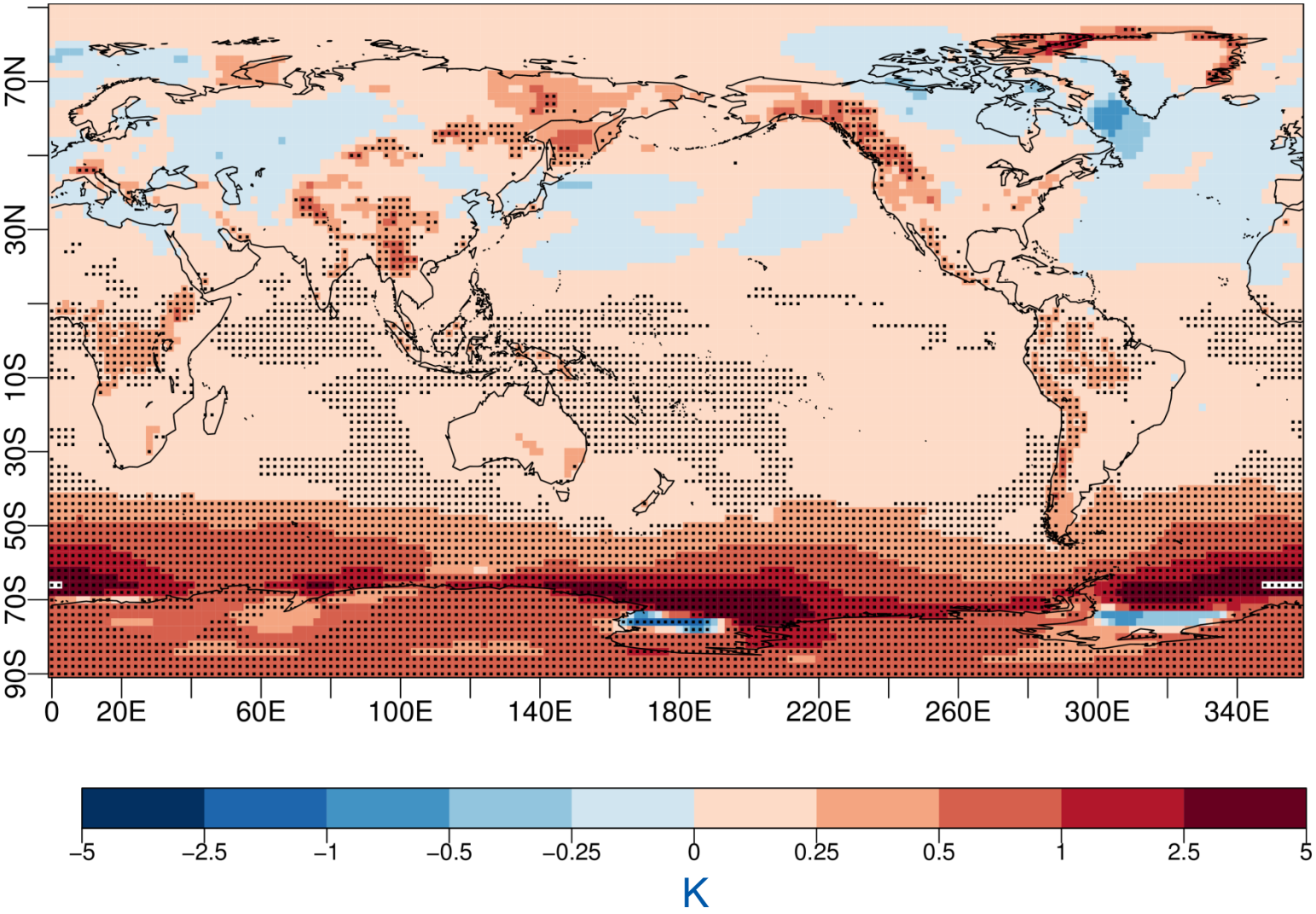
Changing the number of processors only does not affect the results

Difference of mean T2M for MN3-HighProc minus MN3. Stippling = significant at 5%



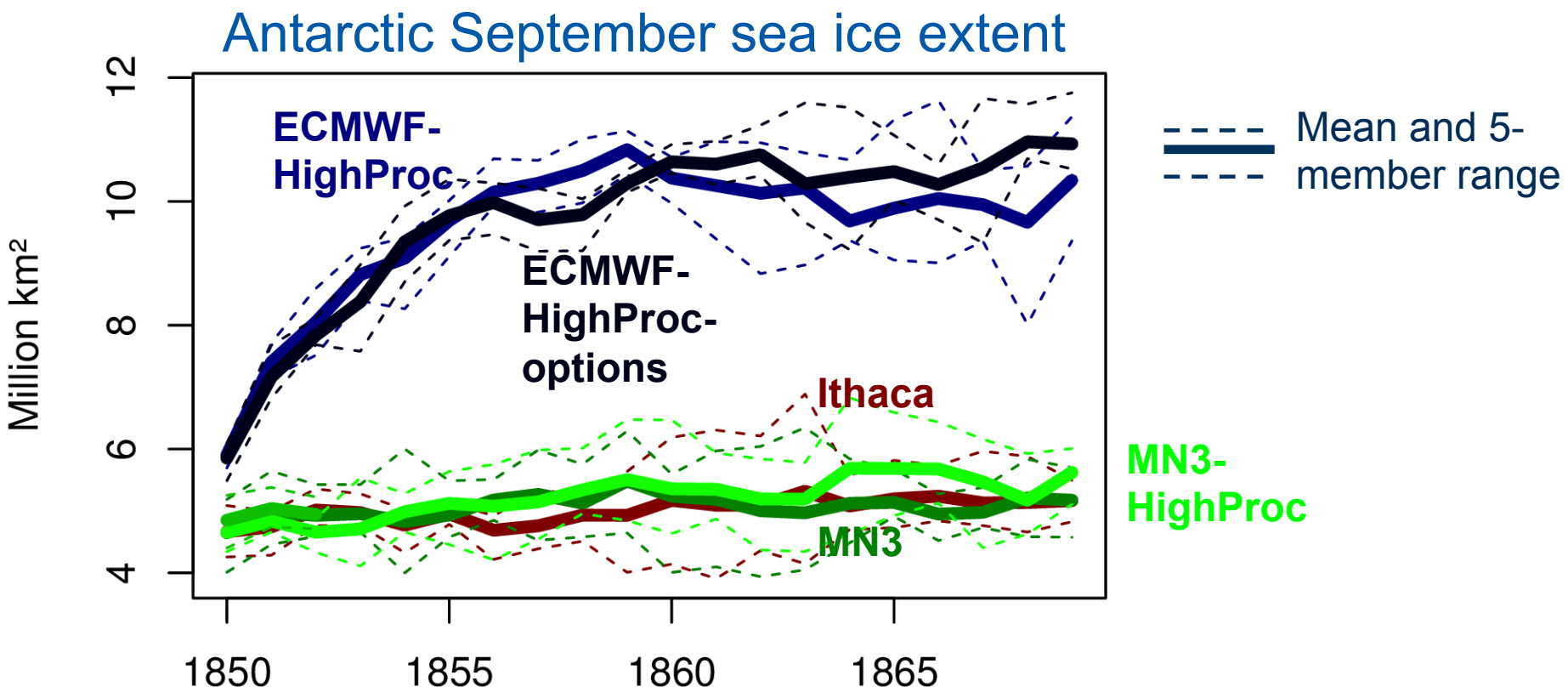
Changing the machine affects the Southern Hemisphere climate in our EC-Earth V3.1 experiments

Difference of mean T2M for MN3-HighProc minus ECMWF-HighProc.
Significant at 5% when stippled.



Machine-dependence of the mean state

ECMWF = CCA machine
MN3 = MareNostrum3 (BSC) machine
Ithaca = Ithaca (IC3) machine



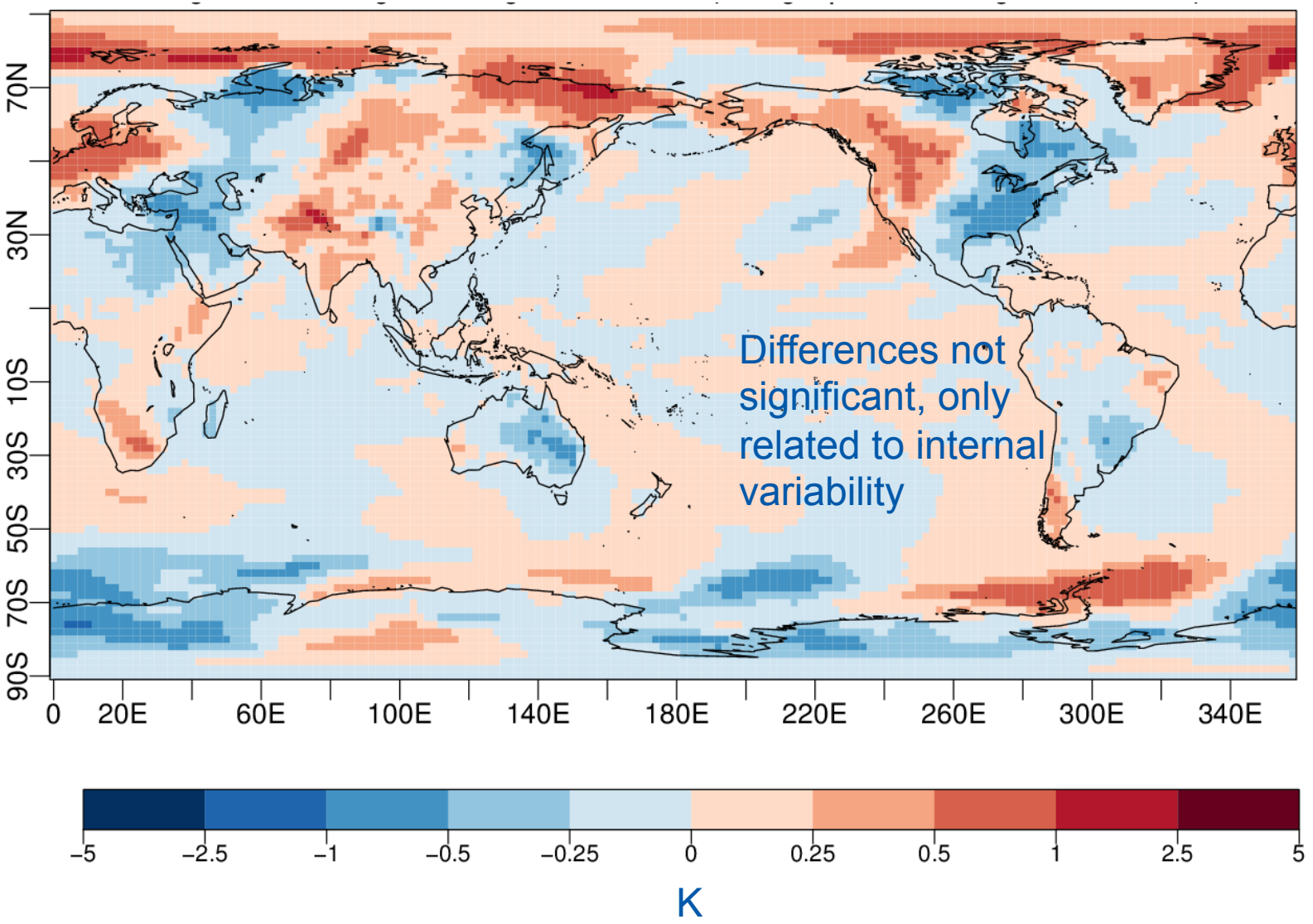
Application to EC-Earth 3.2.beta

- Set of one-year EC-Earth V3.2.beta experiments (To be extended to 20 years).

- Same compilation options used on two different platforms (ECMWF and Marenostrum) -> Compiling with **`-O2 -fp-model precise -fimf-arch-consistency=true -no-fma -g -traceback -vec-report0 -r8 -fpe0`**

Application to EC-Earth 3.2

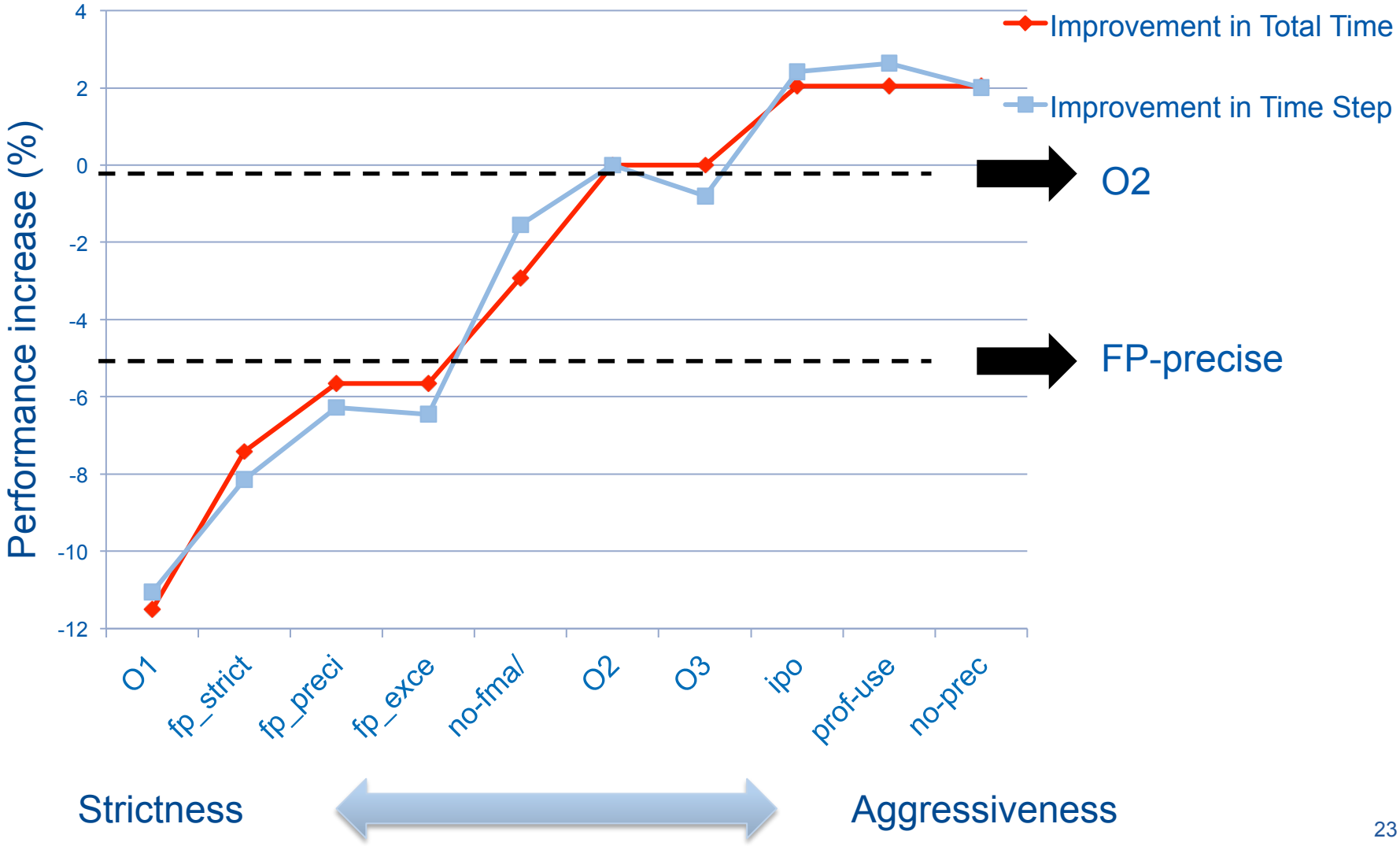
Difference of mean T2M for MN3-HighProc minus ECMWF-HighProc.
Significant at 5% when stippled.



3. Recommendations

Performance/strictness

Evaluation of execution time with short EC-Earth V3.2 runs



Compilations flags

- Best results for an acceptable ratio performance/precision get with the standard flags:

✓ **For development: -fp-model precise –fpe0 -no-fma -O2 -xHost –r8**

✓ **For production: -O2 -xHost –r8**

- More aggressive optimizations (O3, ipo, prof-use) do not improve the performance.

- Stricter FP control does not improve the precision and reduce the performance up to 6%-12%.

- Using approximations for FP operations (no-prec-div/sqrt) does not improve the performance and reduces dramatically the precision and the reproducibility.

Conclusions and outlooks

- We should always check whether climate models are portable from one configuration/platform to another one.
- We simulated the same climate when changing the number of processors and the compilation options in our EC-Earth experiments.
- We get strong climate differences between EC-Earth 3.1 experiments run on different platforms. Running simulations without checking model versions by using control flags (ex: -fpe0) can have dramatic consequencies on the reproducibility.
- We used EC-Earth 3.2.beta compiled with standard flags to produce one-year experiments that were reproducible on a different platform.
- To compile EC-Earth, we suggest the use of the standard flags:
  - ✓ Development: **-fp-model precise –fpe0 -no-fma -O2 -xHost –r8**
  - ✓ Production: **-O2 -xHost –r8**
- **Repeating such a protocol of reproducibility is key to distribute the load of CMIP6 simulations among the EC-Earth partners.**